

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Felipe Werner

Diferença na renda média da população brasileira dentre brancos e não brancos

**Curitiba
2020**

Felipe Werner

Diferença na renda média da população brasileira dentre brancos e não brancos

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Dr. Fernando de Pol Mayer

Curitiba
2020

Diferença na renda média da população brasileira dentre brancos e não brancos

Felipe Werner¹

Prof. Dr. Fernando de Pol Mayer³

Resumo

Neste trabalho foi realizado um estudo para verificar quais fatores possuem correlação com a diferença na renda média de trabalhadores brasileiros, além de avaliar se há desigualdade de salário dentre pessoas com mesmo perfil. O primeiro objetivo foi criar mapas dinâmicos que demonstrem a renda média dos grupos selecionados (brancos e não brancos, homens e mulheres) para todos os municípios do Brasil. O segundo objetivo, foi o ajuste do modelo de Oaxaca-Blinder, este é utilizado para avaliar e quantificar quais fatores interferem na diferença da renda média entre os perfis. Oaxaca-Blinder, Mapas Dinâmicos

Palavras-chave: Oaxaca-Blinder, Mapas Dinâmicos.

Abstract

In this work, a study was developed to verify which factors can influence on the differences in the average income of Brazilian workers. Also, it was assessed whether there is unequal pay among people with the same profile. The first objective was to create a dynamic map that demonstrates the average income of selected groups (white and non-white, men and women) for all municipalities in Brazil. The second objective was to fit the Oaxaca-Blinder model, which is used to assess and quantify which factors interfere in the difference in the average income between the profiles

Keywords: Oaxaca-Blinder, dynamic maps.

1 Introdução

A desigualdade social é um problema que está presente em todos os países do mundo, mas isto é mais grave em países subdesenvolvidos, como é o caso do Brasil. Há várias maneiras de medir a desigualdade, como renda, escolaridade, acesso a serviços básicos, entre outros [1].

Ao longo dos anos, a mídia e empresas de comunicação vem mostrando números de concentração de renda no Brasil, aonde sempre é mostrado que, infelizmente há uma grande desigualdade na distribuição de renda.

Segundo o IPEA (Instituto de Pesquisa Econômica Aplicada), o Brasil possui diferença na renda média dependendo da região do país, gênero, raça, dentre outros fatores [2]. Para avaliar estas informações, ao nível de renda dos trabalhadores, este estudo utilizou a base de dados do CAGED (Cadastro Geral de Empregados e Desempregados). O CAGED foi criado como registro permanente de admissões e dispensa de empregados no Brasil.

O intuito deste estudo foi avaliar a renda média da população brasileira separada em grupos (brancos e não brancos). A partir disso, foi avaliada a diferença na renda média destes grupos e verificou-se quais fatores interferem ou explicam as possíveis diferenças nessas rendas. Para auxílio visual destas diferenças, foi criado um mapa dinâmico, aonde foi plotado a diferença na média salarial entre os grupos.

2 Dados

A base de dados utilizada neste estudo é publica e foi obtida por meio do site do CAGED (Cadastro Geral de Empregados e Desempregados), o período utilizado foi o ano de 2018. O CAGED é um órgão do governo que foi criado como registro permanente de admissões e dispensa de empregados, sob o regime da consolidação das leis de trabalho (CLT). Essa base possui inúmeras informações sobre a população de trabalhadores brasileiros, como por exemplo informações geográficas, como o município de residência o, informações pessoais, gênero e raça além de informações de vínculo empregatício, como salário, local de trabalho, tempo de emprego, dentre outras.

2.1 Resumo dos Dados

Com o intuito de entender os dados, inicialmente foi criado um gráfico contendo a distribuição de renda da população brasileira, conforme mostrado na Figura 1.

Pode-se notar o grande acúmulo de pessoas que recebem até 3.000 reais, em contrapartida, se torna quase imperceptível o volume de pessoas que recebem mais de 7.500 reais.

¹Felipe Werner, werner_felipe@hotmail.com.

³Prof. Dr. Fernando de Pol Mayer - DEST/UFPR.

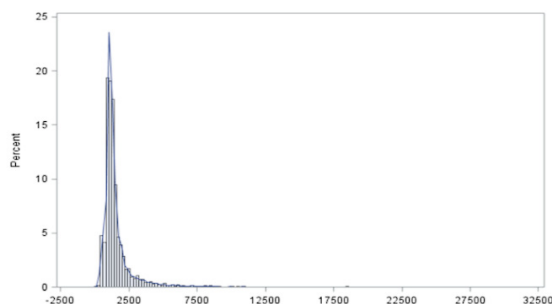


Figura 1: Distribuição de renda população brasileira em 2018

Em um segundo passo, foram gerados quatro grupos, sendo eles, homem branco, homem não branco, mulher branca, mulher não branca. Para os grupos de brancos, foram agrupados os indivíduos brancos e amarelos. No grupo de não brancos foram considerados negros, pardos, indígenas e não identificados. Cabe ressaltar que a informação de raça já vem preenchida na base utilizada.

Avaliando a Figura 2, vemos que no geral, brancos tem média salarial maior do que os não brancos, sendo que homens brancos, grupo de maior média, ganham 35% a mais que as mulheres não brancas.

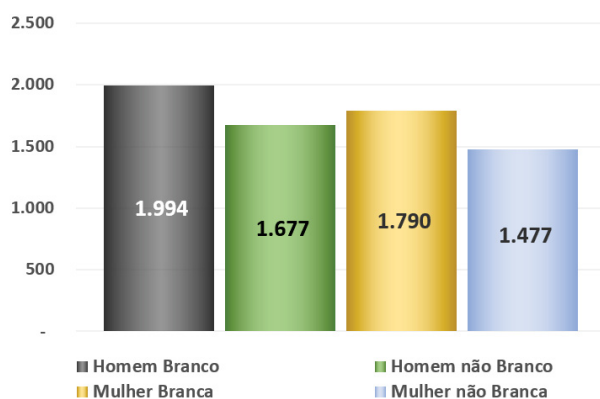


Figura 2: Salário médio por grupo

Na Figura 3, temos as informações de escolaridade, nesta, vemos um comportamento esperado dos dados, aonde a renda aumenta conforme o grau de estudo do indivíduo. Porém, nota-se também que, junto com o aumento na renda média, aumenta também a desigualdade, pois vemos que com o mesmo grau de estudo (mestrado e doutorado por exemplo), homens brancos ganham aproximadamente 57% a mais que mulheres negras.

Na Figura 4 pode-se fazer a mesma análise das informações de escolaridade, aonde vemos que quanto mais velho o indivíduo, maior a renda média.

Quando falamos da desigualdade por região, vemos que São Paulo foi separado de sua região devido ao alto volume de informação, conforme Figura 5. Falando dos resultados, vemos que há uma variação, algumas regiões possuem maiores diferenças na renda média do que as outras, entretanto, há indícios de desigualdade para todas elas.

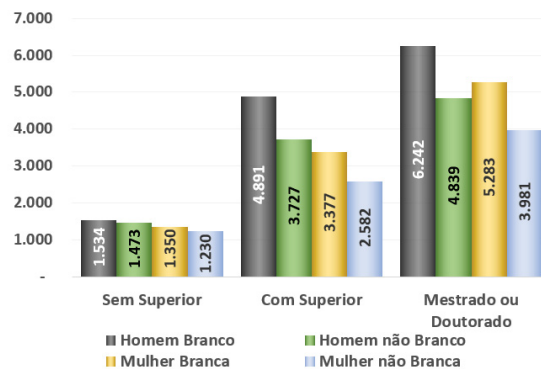


Figura 3: Salário médio por escolaridade e grupo

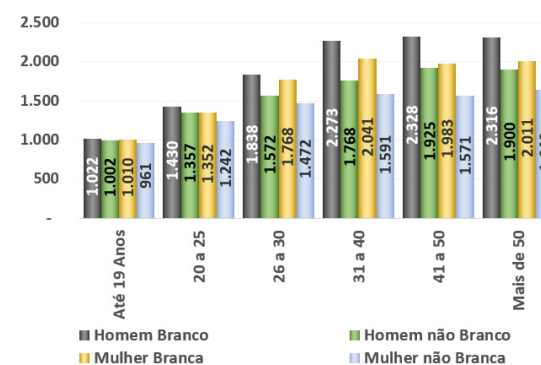


Figura 4: Salário médio por idade e grupo

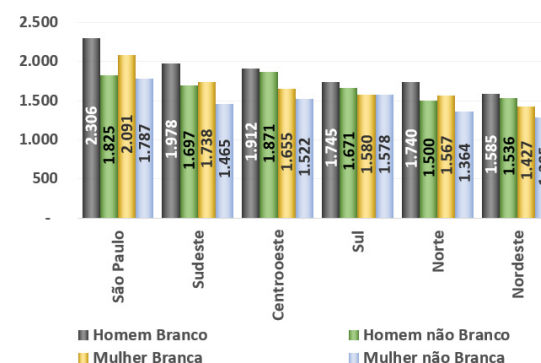


Figura 5: Salário médio por região e grupo

3 Gráficos Dinâmicos

Para uma análise visual mais detalhada e profunda, foi realizado um estudo a nível de município, aonde foram avaliadas as médias de cada cidade do Brasil e plotada em um mapa dinâmico. Para criação de tal gráfico, foi utilizado o pacote **leaflet** [5] do *Software* estatístico R [6].

A Figura 6 representa a diferença salarial entre homens brancos e mulheres negras, aonde quanto mais escura a tonalidade do município, maior a diferença em favor do homem branco. Com base no gráfico, vemos que as maiores desigualdades salariais entre homens brancos e mulheres não brancas estão concentradas nas regiões Centro-Oeste e Norte do país. Conforme mencionado anteriormente, este gráfico é dinâmico, nele também pode ser avaliado a média salarial do homem branco, homem

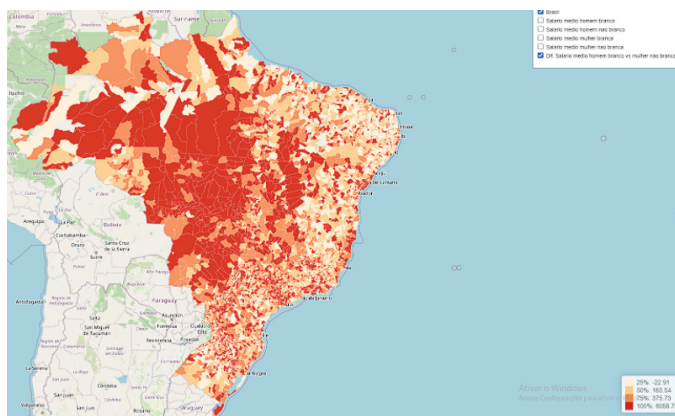


Figura 6: Diferença salarial entre homens brancos e mulheres não brancas

não branco, mulheres brancas e mulheres não brancas separadamente. Para acessar estas informações, utilizar o link: <http://leg.ufpr.br/~fwerner/>.

4 Modelo Aplicado

O modelo aplicado nesse artigo é o denominado de Oaxaca-Blinder [4], que se encontra implementado no pacote **oaxaca** [3] do *software* R. A decomposição de Oaxaca-Blinder consiste em explicar a diferença de uma variável resposta entre dois grupos por meio de uma regressão linear, neste trabalho, foi avaliada a diferença no log dos salários entre brancos (grupo A) e não brancos (grupo B). Para isso, o modelo decompõe esta diferença em dois componentes (*twofold*), sendo eles o "explicável" e o "não explicável". Cada uma destas partes é ajustada considerando um grupo de referência, ou seja, assume-se que um dos grupos é o não discriminado. Para evitar assumir a priori que um dos grupos é não discriminado, foi utilizada uma regressão ponderada, onde não é necessário assumir algum grupo de referência [7]. Em resumo, o modelo consiste em verificar o quanto da diferença média salarial é explicada pelas diferenças entre os preditores nos dois grupos [4].

5 Resultados

Conforme mencionado anteriormente, o intuito deste trabalho é avaliar a diferença salarial entre brancos e não brancos, por isso, foram criados dois grupos de estudo. O grupo A é formado por pessoas das raças "brancas" e "amarelas", e foi chamado de "Brancos". Já o grupo B possui pessoas das raças "preta", "indígena" e "pardo", este foi chamado de "Não Brancos". O modelo foi ajustado considerando o logaritmo do salário como variável resposta. Para identificar fatores que interferem nesta diferença salarial entre os grupos, foram testadas as variáveis idade, região, tempo de emprego, grau de instrução e gênero. A Tabela 1 apresenta a regressão ponderada, sugerida pelo método de Neumark (1988)

[7], entre Brancos e Não Brancos. Pelos resultados mostrados na tabela a seguir, vemos que todas as variáveis são estatisticamente significativas. As variáveis idade e tempo de emprego apresentam estimativas positivas, ou seja, quanto maior o valor, maior a média salarial. Para a variável região, vemos que apenas São Paulo possui coeficiente positivo, sendo assim, é o único a possuir salário maior que a categoria referência (Centro-Oeste). Avaliando o grau de instrução, a categoria referência foi "Com Superior", com isso, vemos que quanto maior o grau de instrução, maior o salário. A última variável analisada foi o gênero, aonde verificamos que o "Feminino" possui média salarial inferior à categoria de referência ("Masculino").

A Tabela 2 apresenta os resultados da decomposição *twofold* do modelo de Oaxaca-Blinder. Esta decomposição tem esse nome pois é dividida em duas partes, a primeira delas é a parte explicável, que é onde a diferença salarial entre brancos e não brancos é explicada segundo as variáveis utilizadas. A segunda é a parte não explicável, que pode ser atribuído à discriminação ou outras variáveis não utilizadas neste estudo. Vale ressaltar que os resultados encontram-se em escala logarítmica. Podemos verificar que o Grupo A possui uma média salarial 0,15 maior quando comparada ao grupo B. Desta diferença, 0,10 pode ser explicada pelas variáveis idade, região, tempo de emprego, grau de instrução e gênero e, 0,05 é a parte não explicada. A parte não explicada da diferença salarial, pode ser causada pela discriminação. Vale ressaltar que a diferença explicada pela variável "Gênero", aonde o salário médio das mulheres é inferior ao dos homens, também deve ser considerada um tipo de discriminação.

A Figura 7 mostra o impacto de cada variável para a diferença explicada (0,10) e não explicada (0,05). Vemos que a Região Nordeste e a escolaridade Superior são as variáveis de maior impacto no componente explicado do modelo, mas a região Norte e o estado de São Paulo também contribuem bastante para o resultado. Por outro lado, vemos que a idade, tempo de emprego e a região sul, não contribuem muito para explicar a diferença, ou seja, possui média salarial mais próxima para os grupos brancos e não brancos.

Por fim, foram avaliados os subcomponentes da parte não explicada da decomposição de Oaxaca-Blinder. Na Figura 8 vemos que o impacto na variável idade é similar para ambos os grupos. Quando avaliamos o tempo de emprego, vemos que o impacto é maior para o grupo "Brancos", isso significa que os ganhos médios são maiores para esse grupo conforme aumenta o tempo de emprego. Outra variável que chama atenção é o "Sem Superior", vemos que ambos os grupos possuem médias negativas, entretanto, o impacto nos "Não Brancos" é maior, isso mostra que para os indivíduos que não possuem grau superior de formação, a média salarial do Grupo "Brancos" é maior.

Tabela 1: Resultado da regressão ponderada para a diferença média de salários entre os grupos “brancos” e “não brancos”.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.7611	0.0024	3253.14	<0.0001
Idade	0.0040	0.0000	98.01	<0.0001
REGIAONORDESTE	-0.2092	0.0017	-122.81	<0.0001
REGIAONORTE	-0.1292	0.0023	-56.62	<0.0001
REGIAOSAO_PAULO	0.0631	0.0017	38.21	<0.0001
REGIAOSUDESTE	-0.0567	0.0017	-33.09	<0.0001
REGIAOSUL	-0.0217	0.0017	-12.47	<0.0001
Tempo_Emprego	0.0001	0.0000	111.13	<0.0001
instrucao_grupMest_Dout	0.4575	0.0054	85.28	<0.0001
instrucao_grupSem_Superior	-0.6927	0.0013	-547.43	<0.0001
sexo_grupMulher	-0.1317	0.0009	-150.72	<0.0001

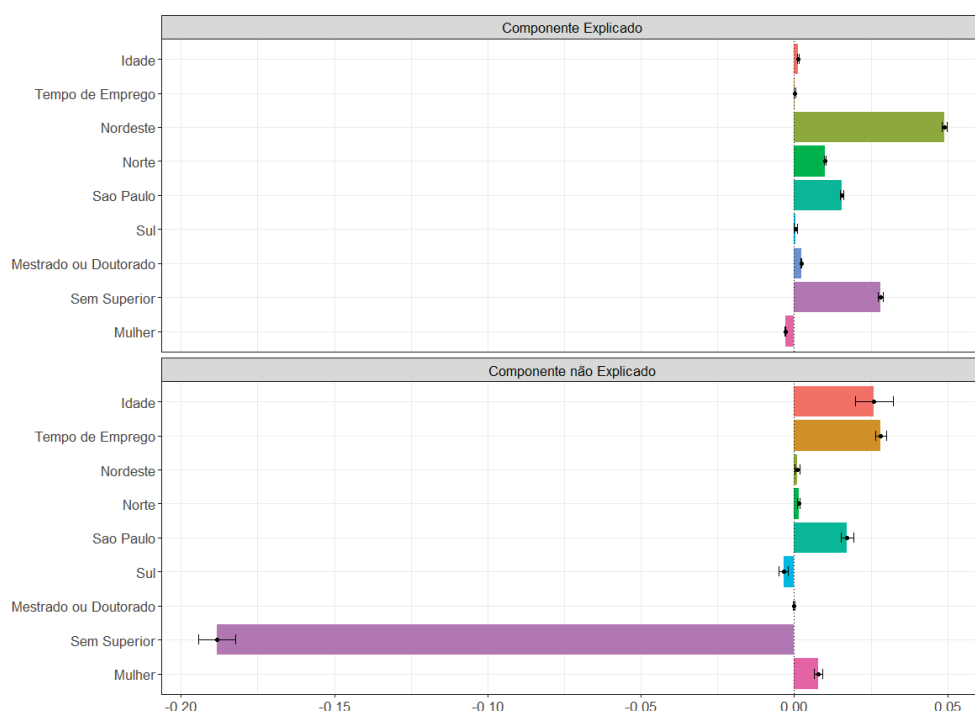


Figura 7: Componentes explicado e não explicado

Tabela 2: Resultado da decomposição *twofold*

Descrição	Valor
Média Salarial do grupo A	7.30
Média Salarial do grupo B	7.15
Diferença salarial	0.15
Componente explicado	0.10
Componente não explicado	0.05

a explicar a diferença no salário médio entre brancos e não brancos, mas não é o suficiente para explicar tudo e que, a parte não explicada provavelmente deve-se ao problema de preconceito ainda existente em nosso país.

6 Conclusão

Pudemos avaliar pelo mapa dinâmico que a diferença salarial entre homens brancos e mulheres não brancas é maior nas regiões centro-oeste e norte do Brasil. Além disso, na decomposição de Oaxaca-Blinder realizada neste trabalho, vimos que algumas covariáveis como Grau de Instrução, Região, Idade e Gênero podem ajudar

Agradecimentos

Agradeço ao Professor Fernando pelo suporte e orientações durante a realização do estudo, aos amigos Jhosefer Silva e Wesley Furriel por todas as discussões e ajudas no desenvolvimento do artigo e a minha esposa Beatriz Lima pelo apoio dado no decorrer de toda especialização.

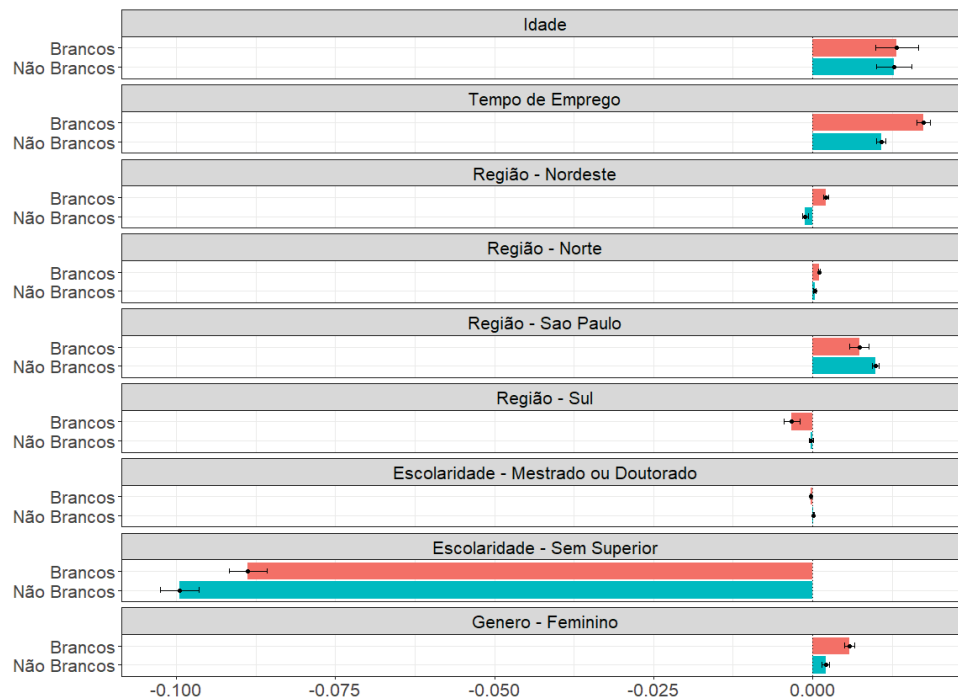


Figura 8: Componente não explicado por variável

Referências

The Journal of Human Resources, University of Wisconsin Press, v. 23, n. 3, p. 279, 1988.

- [1] MENDONÇA, Cláudio. "Países subdesenvolvidos - Características e tentativas de soluções"; Disponível em: <https://educacao.uol.com.br/disciplinas/geografia/paises-subdesenvolvidos-caracteristicas-e-tentativas-de-solucoes.htm>; :text=OutraAcesso em 17 de setembro de 2020
- [2] PORFÍRIO, Francisco. "Desigualdade social"; Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/sociologia/desigualdade-social.htm>. Acesso em 17 de setembro de 2020
- [3] HLAVAC, M. *oaxaca*: Blinder-Oaxaca Decomposition in R. Bratislava, Slovakia, 2018. R package version 0.1.4. Disponível em: <<https://CRAN.R-project.org/package=oaxaca>>.
- [4] JANN, B. A Stata implementation of the Blinder-Oaxaca. *ETH Zurich Sociology Working Paper*, v. 8, n. 5, 2008.
- [5] Joe Cheng, Bhaskar Karambelkar and Yihui Xie (2019). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.3. <https://CRAN.R-project.org/package=leaflet>
- [6] R Development Core Team. (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, url: <http://www.r-project.org>
- [7] NEUMARK, D. Employers' Discriminatory Behavior and the Estimation of Wage Discrimination.